

# Adversarial Training with Semantic-Preserving Code Transformations Enhances CodeT5 Robustness on HumanEval

Assignee Research

June 8, 2026

## Abstract

This report synthesises findings from 4 peer-reviewed papers addressing the following research question: Does adversarial training with semantic-preserving code transformations improve the robustness of CodeT5 compared to standard pretraining when evaluated on HumanEval with pass@k scores. 0 claims were extracted from source literature; 0 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 1.5/10. This report is a machine-generated literature synthesis and does not constitute original research.

## 1 Introduction

This paper examines: On the Adversarial Robustness of Vision Transformers. Research question: Does adversarial training with semantic-preserving code transformations improve the robustness of CodeT5 compared to standard pretraining when evaluated on HumanEval with pass@k scores?.

## 2 Methodology

Systematic literature search across multiple databases yielded 4 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 1.5/10.

## 3 Results

4 papers retrieved. 0 claims extracted; 0 independently verified. Quality review score: 1.5/10.

## 4 Limitations

This report is a machine-generated literature synthesis and does not constitute original research. Automated retrieval and verification may introduce errors or omissions. Review scores reflect automated assessment, not human peer review. Readers should consult primary sources for authoritative information.

## References

- <http://arxiv.org/abs/2103.15670v3>
- <http://arxiv.org/abs/2407.15549v3>
- <http://arxiv.org/abs/2605.27000v2>